

**PEMANFAATAN TEKNIK *STEMMING* UNTUK APLIKASI  
*TEXT PROCESSING* BAHASA INDONESIA**

**SKRIPSI**



Oleh :

**SEPTIAN BAGUS WAHYONO**  
**NPM : 0734010126**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"  
JAWA TIMUR  
2012**

**PEMANFAATAN TEKNIK *STEMMING* UNTUK APLIKASI  
*TEXT PROCESSING* BAHASA INDONESIA**

**SKRIPSI**

**Diajukan Untuk Memenuhi Sebagai Persyaratan  
Dalam Memperoleh Gelar Sarjana Komputer  
Program Studi Teknik Informatika**



**Oleh :**

**SEPTIAN BAGUS WAHYONO**  
**NPM : 0734010126**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"  
JAWA TIMUR  
2012**

## **KATA PENGANTAR**

Puji syukur kehadiran Allah SWT, yang telah berkenan memelihara dan membimbing penulis, sehingga penulis dapat menyelesaikan pengerjaan dan penyusunan laporan ini. Tak lupa, kami ucapkan rasa terimakasih yang teramat dalam kepada semua pihak yang telah membantu penulis untuk menyelesaikan tugas akhir ini. Serta rasa terimakasih kami yang teramat dalam kepada dosen pembimbing Penulis, Bpk. Basuki Rahmat, S.Si, MT dan Ibu Dr. Ir. Ni Ketut Sari, MT yang telah membimbing Penulis dalam menyelesaikan tugas akhir ini.

Disadari bahwa dalam penulisan laporan ini masih jauh dari kata sempurna. Namun, penulis tetap berharap semoga isi dari laporan ini dapat benar-benar berguna baik untuk para penulis khususnya dan para pembaca pada umumnya, maka dari itu kritik dan saran yang membangun sangat kami harapkan.

Surabaya, 20 Februari 2012

Penulis

## UCAPAN TERIMA KASIH

Ucapan terima kasih ini saya persembahkan sebagai perwujudan rasa syukur atas terselesaikannya Laporan Tugas Akhir ini. Ucapan terima kasih ini saya tujukan kepada :

1. Allah SWT, karena berkat Rahmat dan berkahNya kami dapat menyusun dan menyelesaikan Laporan Tugas Akhir ini hingga selesai.
2. Bapak Ir. Sutiyono, MT selaku Dekan Fakultas Teknologi Industri UPN “Veteran” Jawa Timur.
3. Ibu Dr. Ir. Ni Ketut Sari, MT selaku Ketua Jurusan Teknik Informatika UPN “Veteran” Jawa Timur.
4. Bapak Basuki Rahmat, S.Si, MT selaku Dosen Pembimbing Pertama yang telah meluangkan begitu banyak waktu, tenaga, dan pikiran serta sabar dalam membimbing penulis dari awal hingga terselesaikannya Laporan Tugas Akhir ini.
5. Ibu Dr. Ir. Ni Ketut Sari, MT selaku Dosen Pembimbing Kedua yang telah meluangkan begitu banyak waktu, tenaga, dan pikiran serta sabar dalam membimbing penulis dari awal hingga terselesaikannya Laporan Tugas Akhir ini.

6. Kedua orang tua tercinta atas semua doa, dukungan, serta harapan-harapannya pada saat penulis melakukan Tugas Akhir sampai terselesainya penyusunan Laporan Tugas Akhir ini.
7. Adikku Alfian Yuliadi yang selalu memberikan semangat serta dorongan untuk menyelesaikan Tugas Akhir ini.
8. Dian Mega Anjarwati (Bian) yang telah berperan penting dalam memberikan Doa serta Semangat kepada penulis untuk menyelesaikan Tugas Akhir ini.
9. Lucky, Arfin, Eko Sapril, Duwi (Gondrong), Tim Futsal EMJE, Sigit 32, Rizal (Atenk), Bagus (Lek Di) dan semua teman-teman Teknik Informatika yang telah membantu penyusun dalam menyelesaikan Tugas Akhir ini.

Sebagai manusia biasa pasti penulis mempunyai keterbatasan dan banyak sekali kekurangan, terutama dalam penyusunan laporan ini. Oleh karena itu, penulis sangat membutuhkan kritik dan saran yang membangun dalam memperbaiki penulisan laporan ini.

Surabaya, 20 Februari 2012

Tim Penulis

## DAFTAR ISI

	Halaman
KATA PENGANTAR .....	i
UCAPAN TERIMA KASIH.....	ii
DAFTAR ISI.....	iv
DAFTAR GAMBAR .....	vii
DAFTAR TABEL.....	vii
BAB I    PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Perumusan Masalah .....	2
1.3 Batasan Masalah.....	2
1.4 Tujuan .....	3
1.5 Manfaat .....	3
1.6 Metode Penelitian.....	3
1.7 Sistematika Penulisan.....	4
BAB II    TINJAUAN PUSTAKA .....	6
2.1 <i>Information Retrieval Text Processing</i> (Pengolahan Teks) .....	6
2.2 <i>Text Mining</i> .....	10
2.3 <i>Tokenizer</i> .....	11
2.4 <i>Parsing</i> .....	12
2.5 <i>Top Down Parsing</i> .....	12
2.6 <i>Bottom Up Parsing</i> .....	14

2.7	<i>Hybrid Parser</i> .....	15
2.8	<i>Grammar</i> .....	16
2.9	<i>Lexical Analyzer</i> .....	18
2.10	<i>Stoplist</i> .....	19
2.11	Morfologi Bahasa Indonesia .....	20
2.12	<i>Stemming</i> .....	23
2.13	<i>Flow Map</i> .....	26
2.14	<i>Unified Modelling Language</i> .....	27
2.15	Semantik Dalam UML .....	32
2.16	Notasi Dalam UML .....	33
2.16.1	<i>Actor</i> .....	33
2.16.2	<i>Use Case</i> .....	34
2.16.3	<i>Class</i> .....	35
2.16.4	<i>Interface</i> .....	35
2.16.5	<i>Interaction</i> .....	36
2.16.6	<i>Note</i> .....	36
2.16.7	<i>Dependency</i> .....	36
2.16.8	<i>Association</i> .....	37
2.16.9	<i>Generalization</i> .....	37
2.16.10	<i>Realization</i> .....	38
2.17	Embarcadero Delphi 2010 .....	38
BAB III	ANALISIS PERMASALAHAN .....	44
3.1	Analisis Masalah .....	44
3.2	Tahap Pemrosesan Data Teks .....	44

3.3	Algorithma <i>Stemming</i> .....	46
3.3.1	Penjabaran Proses <i>stemming</i> untuk Sebuah Kata ..	48
3.3.2	Peluruhan Awalan ( <i>Prefiks</i> ).....	51
3.3.3	Flowchart Proses <i>Stemming</i> .....	53
BAB IV	PERANCANGAN DAN IMPLEMENTASI.....	55
4.1	Tahap Perancangan .....	55
4.1.1	Perancangan Proses Aplikasi .....	55
4.1.2	Penggunaan Perangkat Lunak.....	55
4.1.3	Penggunaan Perangkat Keras.....	56
4.2	Desain Antar Muka .....	56
4.3	Desain Tabel Daftar Kata Dasar .....	57
4.4	Desain Tabel Daftar <i>Stoplist</i> .....	58
4.5	Implementasi Algorithma Proses <i>Stemming</i> .....	58
BAB V	HASIL PENGUJIAN APLIKASI.....	62
5.1	Tampilan Awal.....	62
5.2	Pengujian Aplikasi .....	63
5.2.1	Pengujian Dengan Input Langsung .....	63
5.2.2	Pengujian Dengan Input File Teks.....	64
BAB VI	PENUTUP.....	72
6.1	Kesimpulan .....	72
6.2	Saran.....	72
	DAFTAR PUSTAKA .....	74



## DAFTAR GAMBAR

	Halaman
Gambar 2.1 Cara kerja <i>Top Down Parser</i> .....	13
Gambar 2.2 Cara kerja <i>Bottom Up Parser</i> .....	14
Gambar 2.3 Cara kerja <i>Hybrid Parser</i> .....	16
Gambar 2.4 Desain dari <i>Porter Stemmer</i> for Bahasa Indonesia .....	26
Gambar 2.5 Metodologi dalam UML .....	29
Gambar 2.6 Notasi <i>Actor</i> .....	33
Gambar 2.7 Notasi <i>Use Case</i> .....	34
Gambar 2.8 Notasi <i>Class</i> .....	35
Gambar 2.9 Notasi <i>Interface</i> .....	35
Gambar 2.10 Notasi <i>Interaction</i> .....	36
Gambar 2.11 Notasi <i>Note</i> .....	36
Gambar 2.12 Notasi <i>Dependency</i> .....	36
Gambar 2.13 Notasi <i>Asociation</i> .....	37
Gambar 2.14 Notasi <i>Generalization</i> .....	37
Gambar 2.15 Notasi <i>Realization</i> .....	38
Gambar 2.16 Tampilan awal delphi.....	40
Gambar 2.17 Daftar komponen Palet Standard delphi .....	41
Gambar 2.18 Daftar komponen Palet <i>Additional</i> .....	41
Gambar 2.19 Daftar komponen Palet <i>Win 32</i> .....	41
Gambar 2.20 Daftar komponen Palet <i>System</i> .....	42
Gambar 2.21 Komponen Palet dialog.....	42
Gambar 2.22 Membuat Aplikasi sederhana dengan Delphi .....	42

Gambar 3.1	Flowchart dari algoritma <i>stemming</i> Bahasa Indonesia.....	53
Gambar 4.1	Perancangan <i>form stemming</i> .....	57
Gambar 4.2	Desain daftar kata dasar .....	57
Gambar 4.3	Desain tabel <i>Stoplist</i> .....	58
Gambar 4.4	Source Code Proses <i>Stemming</i> .....	59
Gambar 4.5	Source Code Proses <i>Stemming</i> imbuhan yang berupa akhiran.....	59
Gambar 4.6	Source Code Proses <i>Stemming</i> akhiran turunan.....	60
Gambar 4.7	Source Code membuka <i>form</i> dialog membaca file bertipe teks .....	61
Gambar 5.1	Tampilan awal <i>form</i> .....	62
Gambar 5.2	Hasil uji aplikasi yang pertama .....	63
Gambar 5.3	Hasil uji aplikasi yang kedua .....	63
Gambar 5.4	Hasil uji aplikasi yang ketiga .....	64
Gambar 5.5	Hasil uji aplikasi yang keempat .....	64
Gambar 5.6	Hasil uji aplikasi yang kelima .....	65
Gambar 5.7	Hasil uji aplikasi yang keenam .....	65
Gambar 5.8	Hasil uji aplikasi yang ketujuh.....	66
Gambar 5.9	Hasil uji aplikasi yang kedelapan.....	66
Gambar 5.10	Hasil uji aplikasi yang kesembilan.....	67
Gambar 5.11	Hasil uji aplikasi yang kesepuluh.....	67

## DAFTAR TABEL

	Halaman
Tabel 2.1      Jenis-jenis file dalam delphi.....	43
Tabel 3.1      Kombinasi awalan dan akhiran yang tidak dibenarkan.....	47
Tabel 3.2      Formula untuk aturan turunan <i>prefiks</i> .....	52
Tabel 4.1      Keterangan tabel kata dasar .....	58
Tabel 4.2      Keterangan tabel <i>stoplist</i> .....	58
Tabel 5.1      Hasil pengujian aplikasi ( Bagian I ).....	68
Tabel 5.2      Hasil pengujian aplikasi ( Bagian II ) .....	69
Tabel 5.3      Hasil pengujian aplikasi ( Bagian III ) .....	70
Tabel 5.4      Hasil pengujian aplikasi ( Bagian IV ) .....	71

**Judul** : PEMANFAATAN TEKNIK *STEMMING* UNTUK  
APLIKASI *TEXT PROCESSING* BAHASA  
INDONESIA

**Dosen pembimbing I** : Basuki Rahmat, S.Si, MT.  
**Dosen Pembimbing II** : Dr. Ir. Ni Ketut Sari, MT.  
**Penulis** : Septian Bagus Wahyono

---

### ABSTRAK

Pencarian informasi berupa dokumen teks atau yang dikenal dengan istilah *Information Retrieval* (IR) merupakan proses pemisahan dokumen-dokumen yang dianggap relevan dari sekumpulan dokumen yang tersedia.

Bertambahnya jumlah dokumen teks yang harus diakses, diikuti dengan meningkatnya kebutuhan pengguna akan perangkat pencarian informasi yang efektif dan efisien. Efektif berarti *user* mendapatkan dokumen yang relevan dengan *query* yang diinputkan. Efisien berarti waktu pencarian yang sesingkat-singkatnya. *Stemming* adalah salah satu cara yang digunakan untuk meningkatkan performa IR dengan cara mentransformasi kata-kata dalam sebuah dokumen teks ke kata dasarnya. Algoritma *Stemming* untuk bahasa yang satu berbeda dengan algoritma *stemming* untuk bahasa lainnya. Sebagai contoh Bahasa Inggris memiliki morfologi yang berbeda dengan Bahasa Indonesia sehingga algoritma *stemming* untuk kedua bahasa tersebut juga berbeda. Proses *stemming* pada teks berbahasa Indonesia lebih rumit dan kompleks karena terdapat variasi imbuhan yang harus dibuang untuk mendapatkan *root word* dari sebuah kata.

Pada Tugas Akhir kali ini membahas tentang penerapan Algoritma *stemming* dalam bentuk aplikasi yang dapat dijalankan di komputer sehingga mempermudah *user* dalam pencarian bentuk dasar suatu kata.

**Kata kunci** : *information retrieval*, temu kembali informasi, *stemming* bahasa indonesia.

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Penggunaan *Data Mining* secara umum diperuntukkan menggali informasi dari data mentah. Salah satu metode yang menarik adalah *Text Mining*, yang digunakan untuk menggali informasi dari data - data dalam bentuk teks seperti buku, makalah, paper, dan lain sebagainya. Kesulitan yang cukup dominan pada *Teks Mining* adalah bagaimana data - data dalam bentuk teks dapat dirubah formatnya agar informasi yang mungkin ada dalam teks-teks tersebut dapat diekstraksi dengan baik.

Dari permasalahan di atas, dikembangkan metode pengelompokan dokumen. Pengelompokan dokumen ini bertujuan untuk mengelompokan dokumen yang merepresentasikan satu topik tertentu. Dengan metode ini, *user* akan menemukan dokumen yang dibutuhkan dengan memilih salah satu kelompok dokumen yang berisi dokumen yang dibutuhkan pengguna tersebut.

Munculnya ide pengelompokan dokumen ini mendorong munculnya berbagai algoritma untuk mengelompokan dokumen. Algoritma-algoritma tersebut diantaranya adalah *K-Means*, *Clustering Based on Frequent Word Sequence* (CFWS), *Naïve Bayes Classifier* (NBC) dan *Minor Component Analysis*.

Sebelum dapat digunakan sebagai data input untuk proses klasifikasi, dokumen harus dilakukan tahapan *preprocessing* yang dikenal dengan istilah

*stemming* yaitu teknik ekstraksi untuk memisahkan kalimat bahasa Indonesia menjadi struktur kata berdasarkan kata dasarnya.

Berdasarkan uraian diatas, maka pada tugas akhir ini berfokus pada pembuatan aplikasi pemrosesan teks bahasa Indonesia sebagai langkah awal dari *text mining*.

## **1.2. Perumusan Masalah**

Rumusan masalah yang digunakan dalam tugas akhir ini adalah :

1. Bagaimana membangun aplikasi yang dapat memisahkan kalimat bahasa Indonesia menjadi beberapa suku kata sesuai dengan kata dasarnya ?
2. Bagaimana menganalisa hasil dari proses pemisahan suku kata tersebut ?

## **1.3. Batasan Masalah**

Agar pembahasan dan penyusunan tugas akhir ini dapat dilakukan secara terarah dan tidak menyimpang serta sesuai dengan apa yang diharapkan, maka perlu ditetapkan batasan-batasan dari masalah yang dihadapi, yaitu :

1. Aplikasi yang dibuat dapat dijalankan pada Sistem Operasi Windows 7.
2. Perangkat lunak yang digunakan untuk membangun aplikasi ini adalah Embarcadero Delphi 2010.
3. Proses *stemming* menggunakan bantuan kamus data dan *algorithm porter for stemmer* untuk bahasa Indonesia.

4. Data yang dijadikan bahan input adalah berupa file dokumen yang berekstensi.txt.
5. Aplikasi yang digunakan hanya untuk pemisahan akhiran dari suatu kata yang telah di inputkan.

#### **1.4. Tujuan**

Tujuan yang ingin dicapai pada pengerjaan tugas akhir ini adalah:

1. Membangun aplikasi yang dapat memisahkan imbuhan berupa akhiran dari kalimat berbahasa Indonesia menjadi struktur kata menurut kata dasarnya.
2. Mengurangi kegagalan proses pemisahan kata pada suatu kalimat.

#### **1.5. Manfaat**

Adapun manfaat dan tujuan yang ingin diperoleh dari pengerjaan tugas akhir ini adalah dapat membuat sistem aplikasi pengolahan teks dengan teknik pemisahan akhiran suatu kata menjadi kata dasarnya, sehingga dapat digunakan untuk kebutuhan pengembangan aplikasi pengolahan kata lainnya.

#### **1.6. Metode Penelitian**

Adapun metode penelitian yang dipergunakan dalam pengerjaan tugas akhir ini adalah :

##### **1. Studi Literatur**

Mencari referensi dan bahan pustaka tentang teori-teori yang berhubungan dengan permasalahan yang akan dikerjakan dalam tugas akhir ini.

## 2. Studi Kasus

Mencari contoh-contoh kasus serupa yang berhubungan dengan permasalahan dalam tugas akhir ini.

### 1.7. Sistematika Penulisan

Sistematika penulisan tugas akhir ini disusun untuk memberikan gambaran umum tentang penelitian yang dijalankan. Sistematika penulisan tugas akhir ini adalah sebagai berikut :

#### **BAB I           PENDAHULUAN**

Bab ini berisi latar belakang masalah, identifikasi masalah, maksud dan tujuan yang ingin dicapai, batasan masalah, metodologi penelitian yang diterapkan dalam memperoleh dan mengumpulkan data, waktu dan tempat penelitian, serta sistematika penulisan.

#### **BAB II          TINJAUAN PUSTAKA**

Membahas berbagai konsep dasar dan teori-teori yang berkaitan dengan topik masalah yang diambil dan hal-hal yang berguna dalam proses analisis permasalahan.

#### **BAB III        ANALISIS PERMASALAHAN**

Menganalisis masalah dari model penelitian untuk memperlihatkan keterkaitan antar variabel yang diteliti serta model matematis untuk analisisnya.

#### **BAB IV        PERANCANGAN DAN IMPLEMENTASI**

Membahas mengenai perancangan dan pengimplementasian aplikasi yang telah dibuat ke perangkat yang akan digunakan.



## **BAB V        PENGUJIAN APLIKASI**

Bab ini membahas pengujian terhadap aplikasi yang telah diimplementasikan tersebut dan hasil pengujiannya.

## **BAB VI        PENUTUP**

Berisi kesimpulan dan saran yang sudah diperoleh dari hasil penulisan tugas akhir.